The URGI plants and bio-agressors genomics annotation system

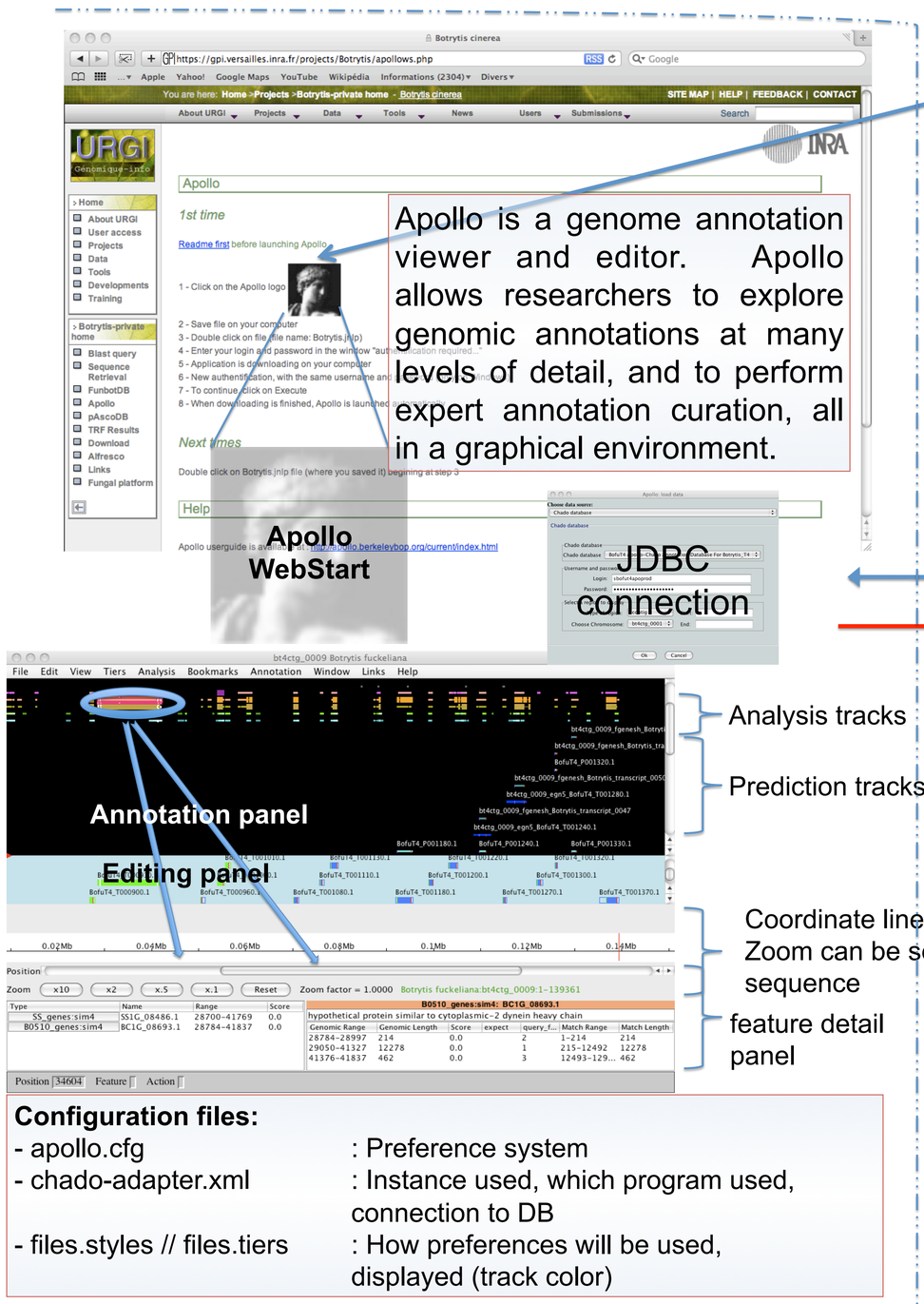
http://urgi.versailles.inra.fr

Baptiste Brault[12], Michael Alaux[1], Fabrice Legeai[3], Sébastien Reboux[1], Isabelle Luyten[1], Stéphanie Sidibé-Bocs[4], Delphine Steinbach[1], Hadi Quesneville[1], Joelle Amselem[12]

[1] INRA URGI Versailles France, [2]INRA BIOGER Versailles France, [3]INRA BiO3P Rennes France, [4]CIRAD UMR DAP Montpellier France

Abstract : The URGI genomic annotation platform, developed in the framework of the GnpAnnot project, relies on well known GMOD tools (http://gmod.org): Apollo, Chado and GBrowse. Apollo is the graphical interface for visualization and annotation edition allowing curators to edit their genes according to evidences (transcript and protein similarity, comparative genomics)**. Manual annotations (gene curation validated/in progress) are saved in a dedicated Chado database and shared at the same time with other community annotations members**. Validated genes/pseudogenes are then committed in the second Chado database accessible by GBrowse.

"Roundtrip" Apollo ⇔ Chado ⇔ GBrowse



Apollo is a genome annotation viewer and editor. Apollo allows researchers to explore genomic annotations at many levels of detail, and to perform expert annotation curation, all in a graphical environment.

Result from the automated annotation have to be carefully verified by manual curation using an editing interface. This way allow to increase in quality and value annotation.
We set up manual genomic annotation platform relying on the international open source Generic Model Organism Database (GMOD http://www.gmod.org) project.
Here is the **data flow management** called "roundtrip" between Chado database, Apollo genome annotation editor and GBrowse genome browser.

Configuration files:
- apollo.cfg : Preference system
- chado-adapter.xml : Instance used, which program used, connection to DB
- files.styles // files.tiers : How preferences will be used, displayed (track color)

Write

Read

Chado4Apollo

Extract genes Curated/Validated

Automated annotation

Gene Prediction
(Fgenesh, Eugene …)
Repeat searching
(TEs, TandemRepeats)
(RepeatMasker, TRF…)
Comparison (ESTs, proteins...)
(Blastn, Blastx, Sim4…)
Comparative genomics ...

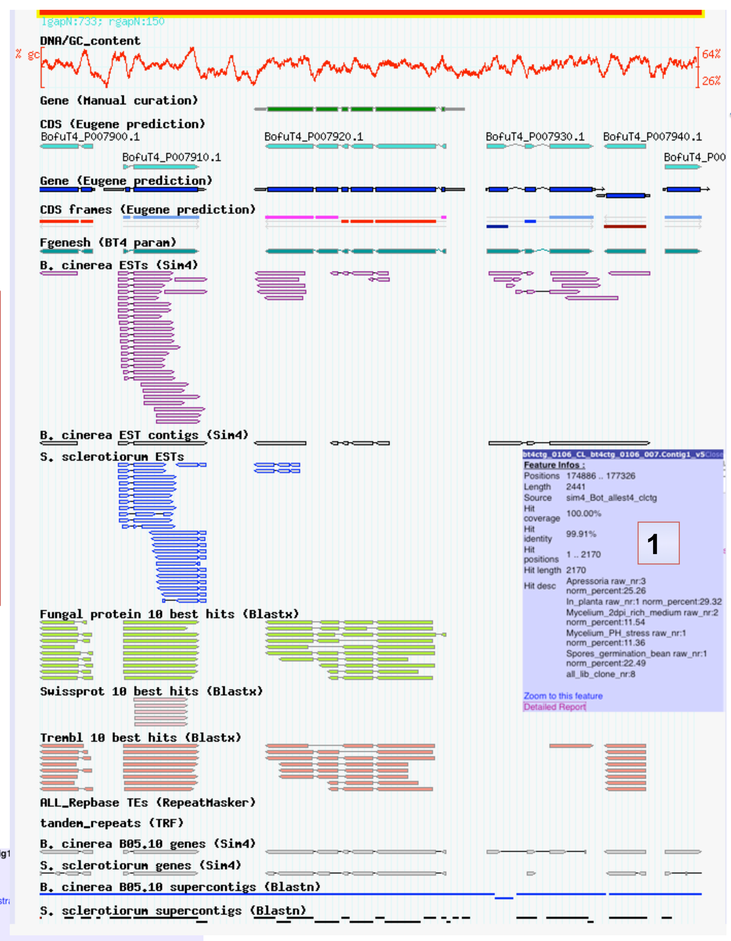
gff3 files

GFF3 files

Update With last curated genes release

Chado4GBrowse

Read



GBrowse graphical interface:

1 : pop up menu
2 : part of detailed report
Both available for each feature

Apollo development:
We are involved in the improvement of Apollo. Some bugs were fixed. New functionalities have been added. They are available in URGI version and expected in next release.
-Owner attribute filled with the username from login on chadoDB
-More information about feature are displayed

Application:
Distributed annotation system has been set up for 2 fungal genomes *Botrytis cineara, Tuber melanosporum* and a plant genome *Grapevine*.
The system for *Leptospheria maculans genome* is in progress.

Perspectives :

Whole tools of the manual curation platform will be improved, and we regularly checked last release of software used (Apollo, GBrowse). We also plan to develop a functional manual annotation system based on the same model.

References:
http://www.gmod.org

Acknowledgements:
Thanks to former URGI team members, GnpAnnot project and GMOD consortium